METHODS AND SYSTEMS FOR UNDERSTANDING A MEANING OF A KNOWLEDGE ITEM USING INFORMATION ASSOCIATED WITH THE KNOWLEDGE ITEM

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application is a continuation-in-part of United States Patent Application Serial No. 09/493,701 filed January 28, 2000 entitled "Meaning-Based Advertising and Relevance Determination," which is is a continuation-in-part of United States Patent No. 6,453,315 filed November 1, 1999 entitled "Meaning-Based Information Organization and Retrieval," which claims priority to United States Provisional Patent Application Serial No. 60/155,667 filed September 22, 1999, all of which are hereby incorporated in their entirety by this reference, and this application claims priority to United States Provisional Patent Application Serial No. 60/491,422 filed July 30, 2003 entitled "Systems and Methods of Organizing and Retrieving Information Based on Meaning," which is hereby incorporated in its entirety by this reference.

FIELD OF THE INVENTION

[0002] The invention generally relates to knowledge items. More particularly, the invention relates to methods and systems for understanding meaning of knowledge items using information associated with the knowledge item.

BACKGROUND OF THE INVENTION

[0003] Two knowledge items are sometimes associated with each other through manual or automated techniques. Knowledge items are anything physical or non-physical that can be represented through symbols and can be, for example, keywords, nodes, categories, people, concepts, products, phrases, documents, and other units of knowledge. Knowledge items can take any form, for example, a single word, a term, a short phrase, a document, or some other structured or unstructured information. Documents include, for example, web pages of various formats, such as HTML, XML, XHTML; Portable Document Format (PDF) files; and word processor and application program document files. For example, a knowledge item, such as, content from a document, can be matched to another knowledge item, such as, a keyword or advertisement. Similarly, a knowledge item, such as, a document, may be associated with another document containing related content so that the two documents can be seen to be related.

[0004] One example of the use of knowledge items is in Internet advertising. Internet advertising can take various forms. For example, a publisher of a website may allow advertising for a fee on its web pages. When the publisher desires to display an advertisement on a web page to a user, a facilitator can provide an advertisement to the publisher to display on the web page. The facilitator can select the advertisement by a variety of factors, such as demographic information about the user, the category of the web page, for example, sports or entertainment, or the content of the web page. The facilitator can also match the content of the web page to

a knowledge item, such as a keyword, from a list of keywords. An advertisement associated with the matched keyword can then be displayed on the web page. A user may manipulate a mouse or another input device and "click" on the advertisement to view a web page on the advertiser's website that offers goods or services for sale.

[0005] In another example of Internet advertising, the actual matched keywords are displayed on a publisher's web page in a Related Links or similar section. Similar to the example above, the content of the web page is matched to the one or more keywords, which are then displayed in the Related Links section, for example. When a user clicks on a particular keyword, the user can be directed to a search results page that may contain a mixture of advertisements and regular search results. Advertisers bid on the keyword to have their advertisements appear on such a search results page for the keyword. A user may manipulate a mouse or another input device and "click" on the advertisement to view a web page on the advertiser's website that offers goods or services for sale.

[0006] Advertisers desire that the content of the web page closely relate to the advertisement, because a user viewing the web page is more likely to click on the advertisement and purchase the goods or services being offered if they are highly relevant to what the user is reading on the web page. The publisher of the web page also wants the content of the advertisement to match the content of the web page, because the publisher is often compensated if the user clicks on the advertisement and a mismatch could be offensive to either the advertiser or the publisher in the case of sensitive content.

[0007] Various methods have been used to match keywords with content. Most of these methods have involved a form of text matching, for example, matching the keywords with words contained in the content. The problem with text matching is that words can relate to multiple concepts, which can lead to mismatching of content to keyword.

[0008] For example the term "apple" can relate to at least two concepts. Apple can refer to the fruit or the computer company by the same name. For example, a web page can contain a news story about Apple Computer and the most frequently used keyword on the web page, in this case "apple", could be chosen to represent the web page. In this example, it is desirable to display an advertisement relating to Apple Computer and not apple, the fruit. However, if the highest bidder on the keyword "apple" is a seller of apples and if the keyword "apple" is matched to the web page, the advertisement about apples, the fruit, would be displayed on the web page dealing with Apple, the computer company. This is undesirable, because a reader of the web page about a computer company is likely not also interested in purchasing apples.

[0009] Mismatching of knowledge items, such as keywords, to content can result in irrelevant advertisements being displayed for content. It is, therefore, desirable to understand the meaning of knowledge items.

SUMMARY

[0010] Embodiments of the present invention comprise systems and methods that understand the meaning of knowledge items using related information. One aspect of an embodiment of the present invention comprises receiving a knowledge item and receiving related information associated with the knowledge item. Such related information may include a variety of information, such as, related documents and related data. Another aspect of an embodiment of the present invention comprises determining at least one related meaning based on the related information and determining a meaning for the knowledge item based at least in part on the related meaning of the related information. A variety of algorithms using the related meaning may be applied in such systems and methods. Additional aspects of the present invention are directed to computer systems and computer-readable media having features relating to the foregoing aspects.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] These and other features, aspects, and advantages of the present invention are better understood when the following Detailed Description is read with reference to the accompanying drawings, wherein:

[0012] FIG. 1 illustrates a block diagram of a system in accordance with one embodiment of the present invention;

[0013] FIG. 2 illustrates a flow diagram of a method in accordance with one embodiment of the present invention; and

[0014] FIG. 3 illustrates a flow diagram of a subroutine of the method shown in FIG. 2.

DETAILED DESCRIPTION OF SPECIFIC EMBODIMENTS

[0015] The present invention comprises methods and systems for understanding the meaning of knowledge items using the knowledge item itself as well as information associated with the knowledge item. Reference will now be made in detail to exemplary embodiments of the invention as illustrated in the text and accompanying drawings. The same reference numbers are used throughout the drawings and the following description to refer to the same or like parts.

[0016] Various systems in accordance with the present invention may be constructed. FIG. 1 is a diagram illustrating an exemplary system in which exemplary embodiments of the present invention may operate. The present invention may operate, and be embodied in, other systems as well.

[0017] The system 100 shown in FIG. 1 includes multiple client devices 102a-n, server devices 104, 140 and a network 106. The network 106 shown includes the Internet. In other embodiments, other networks, such as an intranet may be used. Moreover, methods according to the present invention may operate in a single computer. The client devices 102a-n shown each include a computer-readable medium, such as a random access memory (RAM) 108, in the embodiment shown coupled to a processor 110. The processor 110 executes a set of computer-executable program instructions stored in memory 108. Such processors may include a

microprocessor, an ASIC, and state machines. Such processors include, or may be in communication with, media, for example computer-readable media, which stores instructions that, when executed by the processor, cause the processor to perform the steps described herein. Embodiments of computer-readable media include, but are not limited to, an electronic, optical, magnetic, or other storage or transmission device capable of providing a processor, such as the processor in communication with a touch-sensitive input device, with computer-readable instructions. Other examples of suitable media include, but are not limited to, a floppy disk, CD-ROM, magnetic disk, memory chip, ROM, RAM, an ASIC, a configured processor, all optical media, all magnetic tape or other magnetic media, or any other medium from which a computer processor can read instructions. Also, various other forms of computer-readable media may transmit or carry instructions to a computer, including a router, private or public network, or other transmission device or channel, both wired and wireless. The instructions may comprise code from any computer-programming language, including, for example, C, C++, C#, Visual Basic, Java, and JavaScript.

[0018] Client devices 102a-n may also include a number of external or internal devices such as a mouse, a CD-ROM, a keyboard, a display, or other input or output devices. Examples of client devices 102a-n are personal computers, digital assistants, personal digital assistants, cellular phones, mobile phones, smart phones, pagers, digital tablets, laptop computers, a processor-based device and similar types of systems and devices. In general, a client device 102a-n may be any type of processor-based platform connected to a network 106 and that interacts with one or more

application programs. The client devices 102a-n shown include personal computers executing a browser application program such as Internet Explorer™, version 6.0 from Microsoft Corporation, Netscape Navigator™, version 7.1 from Netscape Communications Corporation, and Safari™, version 1.0 from Apple Computer. Through the client devices 102a-n, users 112a-n can communicate over the network 106 with each other and with other systems and devices coupled to the network 106.

[0019] As shown in FIG. 1, server devices 104, 140 are also coupled to the network 106. The server device 104 shown includes a server executing a knowledge item engine application program. The server device 140 shown includes a server executing a content engine application program. Similar to the client devices 102a-n, the server devices 104, 140 shown each include a processor 116, 142 coupled to a computer readable memory 118, 144. Server devices 104, 140 are depicted as a single computer system, but may be implemented as a network of computer processors. Examples of server devices 104, 140 are servers, mainframe computers, networked computers, a processor-based device and similar types of systems and devices. Client processors 110 and server processors 116, 142 can be any of a number of well known computer processors, such as processors from Intel Corporation of Santa Clara, California and Motorola Corporation of Schaumburg, Illinois.

[0020] Memory 118 of the server device 104 contains a knowledge item processor application program, also known as a knowledge item processor 124. The knowledge item processor 124 determines a meaning for knowledge items. Meaning

can be a representation of context and can be, for example, a vector of weighed concepts or groups or clusters of words. The knowledge items can be received from other devices connected to the network 106, such as, for example, the server device 140.

[0021] The knowledge item processor 124 may also match a knowledge item, such as a keyword, to an article, such as, a web page, located on another device connected to the network 106. Articles include, documents, for example, web pages of various formats, such as, HTML, XML, XHTML, Portable Document Format (PDF) files, and word processor, database, and application program document files, audio, video, or any other information of any type whatsoever made available on a network (such as the Internet), a personal computer, or other computing or storage means. The embodiments described herein are described generally in relation to documents, but embodiments may operate on any type of article. Knowledge items are anything physical or non-physical that can be represented through symbols and can be, for example, keywords, nodes, categories, people, concepts, products, phrases, documents, and other units of knowledge. Knowledge items can take any form, for example, a single word, a term, a short phrase, a document, or some other structured or unstructured information. The embodiments described herein are described generally in relation to keywords, but embodiments may operate on any type of knowledge item.

[0022] Memory 144 of server device 140 contains a content engine application program, also known as a content engine 146. In one embodiment, the content engine

146 receives a matched keyword from the knowledge item engine 124 and associates a document, such as an advertisement, with it. The advertisement is then sent to a requester's website and placed in a frame on a web page, for example. In one embodiment, the content engine 146 receives requests and returns content, such as advertisements, and matching is performed by another device.

[0023] The knowledge item engine 124 shown includes an information locator 134, an information processor 136, a knowledge item processor 135 and a meaning processor 136. In the embodiment shown, each comprises computer code residing in the memory 118. The knowledge item processor 135 receives a keyword and identifies known information about the keyword. The known information may include, for example, one or more concepts associated with one or more terms parsed from the keyword. A concept can be defined using a cluster or set of words or terms associated with it, where the words or terms can be, for example, synonyms. For example, the term 'apple' may have two concepts associated with it – fruit and computer company – and thus, each may have a cluster or set of related words or terms. A concept can also be defined by various other information, such as, for example, relationships to related concepts, the strength of relationships to related concepts, parts of speech, common usage, frequency of usage, the breadth of the concept and other statistics about concept usage in language.

[0024] The information locator 134 identifies and retrieves related information associated with keywords. In the embodiment shown, the related information could include related documents and additional related data. The related documents could

include the text of the advertisements and the destination web site from advertisers that have bid on a keyword. The additional related data could include other keywords purchased by the advertisers, search results on a keyword from a search engine, cost per click data on the advertisers, and data related to the success rate of the advertisements. Some of this information can be obtained, for example, from the server device 140. The information processor 136 processes the related information located by the information locator 134 to determine at least one related meaning for the located related information. This related meaning and the known information about the keyword are then passed to the meaning processor 137. The meaning processor 137 uses the known information about the keyword and the related meaning to determine the meaning of the keyword. Note that other functions and characteristics of the information locator 134, knowledge item processor 135, information processor 136, and meaning processor 137 are further described below.

[0025] Server device 104 also provides access to other storage elements, such as a knowledge item storage element, in the example shown a knowledge item database 120. The knowledge item database can be used to store knowledge items, such as keywords, and their associated meanings. Server device 140 also provides access to other storage elements, such as a content storage element, in the example shown a content database 148. The content database can be used to store information related to knowledge items, for example documents and other data related to knowledge items. Data storage elements may include any one or combination of methods for storing data, including without limitation, arrays, hashtables, lists, and

pairs. Other similar types of data storage devices can be accessed by the server device 104.

[0026] It should be noted that the present invention may comprise systems having different architecture than that which is shown in FIG. 1. For example, in some systems according to the present invention, the information locator 134 may not be part of the knowledge item engine 124, and may carry out its operations offline. The system 100 shown in FIG. 1 is merely exemplary, and is used to explain the exemplary methods shown in FIGS. 2-3.

[0027] Various methods in accordance with the present invention may be carried out. One exemplary method according to the present invention comprises receiving a knowledge item, receiving related information associated with the knowledge item, determining at least one related meaning based on the related information, and determining a knowledge item meaning for the knowledge item based at least in part on the related meaning of the related information. The related information may be associated with the knowledge item in any way, and determined to be related in any way. The related information may comprise related articles and related data. Some examples of related articles comprise an advertisement from an advertiser who has bid on a knowledge item and a web page associated with the advertisement. The knowledge item can be, for example, a keyword. An example of related data comprises cost per click data and success rate data associated with the advertisement. In one embodiment, the knowledge item meaning may comprise a weighted vector of concepts or related clusters of words.

[0028] In one embodiment, the knowledge item is processed after it is received to determine any known associated concepts. A concept can be defined by a cluster or group of words or terms. A concept can further be defined by various other information, such as, for example, relationships to related concepts, the strength of relationships to related concepts, parts of speech, common usage, frequency of usage, the breadth of the concept and other statistics about concept usage in language. In one embodiment, determining the knowledge item meaning comprises determining which of the associated concepts represents the knowledge item meaning.

[0029] In one embodiment, the knowledge item comprises a plurality of concepts and the related meaning comprises a plurality of concepts and determining the knowledge item meaning comprises establishing a probability for each knowledge item concept that the knowledge item should be resolved in part to the knowledge item concept, determining a strength of relationship between each knowledge item concept and each related meaning concept, and adjusting the probability for each knowledge item concept based on the strengths. In one embodiment, the knowledge item has a plurality of concepts and a plurality of related meanings are determined, where each related meaning has a plurality of concepts. A knowledge item meaning determination involves establishing a probability for each knowledge item concept that the knowledge item should be resolved in part to the knowledge item concept and establishing a probability for each related meaning concept that the knowledge item should be resolved in part to the related meaning concept.

[0030] FIGs. 2-3 illustrate an exemplary method 200 in accordance with the present invention in detail. This exemplary method is provided by way of example, as there are a variety of ways to carry out methods according to the present invention. The method 200 shown in FIG. 2 can be executed or otherwise performed by any of various systems. The method 200 is described below as carried out by the system 100 shown in FIG. 1 by way of example, and various elements of the system 100 are referenced in explaining the example method of FIGs. 2-3. The method 200 shown provides an understanding of the meaning of a keyword using information associated with the keyword.

[0031] Each block shown in FIGs. 2-3 represents one or more steps carried out in the exemplary method 200. Referring to FIG. 2, in block 202, the example method 200 begins. Block 202 is followed by block 204 in which a keyword is received by the knowledge item engine 124. The keyword can for example, be received from an external database through network 106, such as the content database 148 or can be received from other sources.

[0032] Next in block 206, the keyword is processed by knowledge item processor 135 to determine known information about the keyword. For example, the keyword may have one or more concepts associated with it. Each concept may have an associated cluster or group of words. A concept can also be defined by various other information, such as, for example, relationships to related concepts, the strength of relationships to related concepts, parts of speech, common usage, frequency of usage, the breadth of the concept and other statistics about concept usage in language.

[0033] For example, for the term apple there may be two possible associated concepts. The first concept of apple the fruit can be defined with relationships to related words or concepts, such as, fruit, food, pie, and eat. The second concept of apple the computer company can be defined with relationships to related words or concepts, such as, computer, PC, and technology. A keyword can be a short phrase, in which case, the phrase can be broken down by the knowledge item processor 135, for example, into individual terms. In such example, the knowledge item processor 135 can further determine concepts associated with each term. In some embodiments, the keyword will not have any information associated with it.

[0034] Block 206 is followed by block 208 in which related information associated with the keyword is identified by the information locator 134 and received by the information processor 136. The related information can include documents, such as, the text of advertisements and destination websites from advertisers who have bid on a keyword, web search results on the keyword itself, and related data, such as, other keywords bid on by the advertisers, the cost per click that the advertisers associated with the keyword are paying, the number of times a user has bought an item after clicking through an associated advertisement to an advertiser's website. This related information can be located from a variety of sources, such as, for example, the server device 140, the advertiser's websites, and search engines.

[0035] Block 208 is followed by block 210, in which the at least one related meaning is determined from the related information by the information processor 136. For example, for each individual related document a meaning could be determined or

an overall meaning for all of the documents could be determined. For example, if the documents include the text of five advertisements associated with the keyword, a related meaning for each advertisement could be determined or the meanings of all five advertisements could be combined to provide an overall related meaning. In one embodiment, documents are processed to determine a vector of weighted concepts contained in the documents. The vector of weighted concepts can represent the meaning of the document. For example, if the advertisement relates to selling Apple Computers, the meaning of such an advertisement may be fifty percent computers, thirty percent Apple Computers and twenty percent sales. The related data can be used, for example, to adjust the weights of the meanings of individual documents or of the overall related meaning. Alternatively, the meaning of a document could be related clusters of words.

[0036] Block 210 is followed by block 212, in which the meaning of the keyword is determined based on the related meaning or meanings by meaning processor 137. Meaning processor 137 receives the related meaning or meanings from information processor 136 and the processed keyword from knowledge item processor 135. For example, in block 212, the meaning processor would receive the keyword apple and its related two concepts from the knowledge item processor and would receive the related meaning of the advertisement for Apple Computers from the information processor 136. A variety of methods could be used to determine the meaning of the keyword based on the related meaning or meanings received from the information processor 136. For example, the related meaning can be used as a clue to

determine the best concept to associate with the keyword to provide a meaning for the keyword. Where the related meaning is, for example, fifty percent computer, thirty percent Apple Computers and twenty percent sales the relationship between the weighted concepts of the related meaning and the concepts of the keyword could be used to indicate that the keyword apple should be associated with the concept of the computer company. Alternatively, the related meaning or meanings and related data can be used to develop a new meaning for the keyword.

[0037] Any one or more of a variety of related information may be used to determine the meaning of a keyword. The examples of related information that may be used to determine the meaning of a keyword include, without limitation, one or more of the following:

- The text of advertisements associated with advertisers who have currently bid on the knowledge item.
- The destination web page or web pages for the advertisements.
- Text of advertisements from advertisers who have in the past bid on the keyword.
- Other keywords bid on by the advertisers who currently have bid on the keyword.
- Search results on the keyword from a search engine.
- The number of people who have bought an item, after viewing the advertisement, from an advertiser's website that is associated with the keyword.

[0038] There are a variety of other related information that may be included, and these are only examples. Moreover, this related information may be given different weights depending on some of the information. For example, the text of advertisements of current advertisers may be weighted more than the text of advertisements of former advertisers associated with the keyword. Further, the items associated with the advertiser with the highest cost per click may be weighted more based on the cost per click.

[0039] FIG. 3 illustrates an example of a subroutine 212 for carrying out the method 200 shown in FIG. 2. The subroutine 212 determines the meaning of the keyword using a related meaning or related meanings. An example of subroutine 212 is as follows.

[0040] The subroutine begins at block 300. At block 300, probabilities for each set of words associated with the keyword are established. For example, in one embodiment each keyword can comprise one or more terms and each term can have one or more concepts associated with it. For purposes of this example, the keyword comprises a single term with at least two related concepts. In block 300, each concept associated with the keyword is given an *a priori* probability of the keyword being resolved to it. This *a priori* probability can be based on information contained in a network of interconnected concepts and/or on previously collected data on the frequency of each term being resolved to the concept.

[0041] Block 300 is followed by block 302, in which the strength of the relationship is determined between the keyword concepts and the related meaning or

meanings concepts. For example, in one embodiment the related meaning may be comprised of a weighed set of concepts. A strength is determined for the relationship between each keyword concept and each related meaning concept. The weight of each related meaning concept can be used to adjust the strength of the relationship between the related meaning concepts and the keyword concept. The strength can reflect the probability of co-occurrence between concepts, or some measure of closeness of the two concepts, which can be derived from ontological data.

[0042] Block 302 is followed by block 304, in which the strengths computed in block 302 are used to adjust the probability of the keyword being resolved to each of its associated concepts. For example, the strengths determined for the relationship between each keyword concept and each related meaning concept are used to adjust the probability of each keyword concept being considered. In one embodiment, after the probabilities for the keyword concepts have been adjusted, the probabilities are normalized to one. The steps occurring in blocks 302 and 304 can be repeated a number of times to boost the impact of the strengths of the relationships on the probabilities.

[0043] In one embodiment, the keyword can comprise multiple concepts and multiple related meanings may each comprise multiple concepts. In this embodiment, the keyword meaning can be determined by establishing a probability for each keyword concept that the keyword should be resolved in part to the keyword concept and a probability for each related meaning concept that the keyword should be

resolved in part to the related meaning concept. These probabilities can be established in the manner described above with respect to FIG. 3.

[0044] Returning now to FIG. 2, block 212 is followed by block 214 in which the meaning of the keyword is associated with the keyword and stored. The keyword and its associated meaning could be stored together, for example, in the knowledge item database 120, or could be stored separately in separate databases.

[0045] While the above description contains many specifics, these specifics should not be construed as limitations on the scope of the invention, but merely as exemplifications of the disclosed embodiments. Those skilled in the art will envision many other possible variations that are within the scope of the invention.